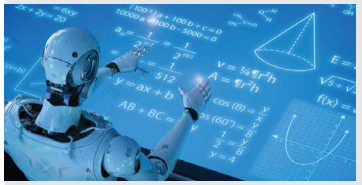
# 22EE201 PROBABILITY THEORY AND STATISTICS FOR MACHINE LEARNING



Source: https://gnindia.dronacharya.info/blog/hashtag/probability-theory-and-statistics/

Hours Per Week :

| L | T | P | C |
|---|---|---|---|
| 3 | 2 | 0 | 4 |

**PREREQUISITE KNOWLEDGE:** Basic knowledge in statistics and mathematics.

**COURSE DESCRIPTION AND OBJECTIVES:**

This course is mainly deals with the essential components of Machine learning i.e probability theory and statistics. It also provides students with foundation in statistics and probability such as descriptive statistics, correlation, probability, random variables, correlation, and regression. The course emphasizes statistics to solve engineering and management problems.

## MODULE-1

**UNIT-1** **12L+8T+0P=20 Hours**

**INTRODUCTION TO MACHINE LEARNING, BASIC PROBABILITY AND RANDOM VARIABLES:**

**Introduction to Machine Learning:** Human learning and its types, Machine learning and its types, Applications of machine learning, Machine Learning activities, Data types in Machine Learning, structure of data, Central tendency, data spread, Measuring data value position.

**Basic Probability:** Random Experiments, The Concept of Probability, Axioms of Probability, Conditional Probability, Theorems on Probability, Conditional Probability, Independent Events, Bayes' Theorem, Combinatorial Analysis.

**Random Variables:** Random variables, Types of random variables, jointly distributed random variables, Independent random variables, Conditional distributions.

**UNIT-2** **12L+8T+0P=20 Hours**

**EXPECTATION AND PROBABILITY DISTRIBUTIONS:**

**Expectation:** Expectation, expected value and variance of sums of random variables, Standard Deviation, Covariance and variance of sums of random variables, Moment generating functions, Chebyshev's inequality and the weak law of large numbers, Other measures of Central Tendency, Percentiles, Measures of Dispersion, Skewness and Kurtosis.

**Probability Distributions:** The Binomial Distribution and properties, The Law of Large Numbers for Bernoulli Trials, Normal Distribution and properties, Poisson Distribution and properties, Relationships among Binomial, Normal and Poisson Distributions, Central Limit Theorem. Multinomial, Hypergeometric, Uniform, Cauchy, Gamma, Beta.

**PRACTICES:**

- Compute different statistical measures for a given data.
- Analyse the sample data to accept or reject statements regarding population parameters.
- Identify the situations we can apply binomial, Poisson, and Normal distributions.

**SKILLS:**

- ✓ *Collect the data from newspapers and present it graphically*
- ✓ *Evaluate the various measures of central tendency and dispersion data collected from the various sources*
- ✓ *Analyse the data using measures of central tendency.*

## MODULE - 2

**UNIT-1** **12L+8T+0P=20 Hours**

**SAMPLING THEORY:**

Population and Sample, Statistical Inference, Random Samples, Sample Statistics Sampling Distributions, Sample Mean, Sampling Distributions - Means, Proportions, Differences and Sums, Sample Variance, Sampling Distribution of Variances, Sampling Distribution of Ratios of Variances, Chi-Square, Student's t, and F Distributions, Relationships among Chi-Square, t, and F Distributions. Frequency Distributions, Relative Frequency Distributions Computation of Mean, Variance, and Moments for Grouped Data.

**UNIT-2** **12L+8T+0P=20 Hours**

**CURVE FITTING, REGRESSION AND CORRELATION:**

Curve Fitting, Regression, Method of Least Squares, Least-Squares Line, The Least-Squares Parabola, Multiple Regression, Standard Error of Estimate, Linear Correlation Coefficient, Generalized Correlation Coefficient, Rank Correlation, Probability Interpretation of Regression and Correlation, Sampling Theory of Regression and Correlation, Correlation and Dependence.

**PRACTICES:**

- Fit an appropriate curve for a given set of data.
- Multiple linear regression with two independent variables.
- Developing multiple regression model using excel.
- Conducting Correlation and regression analysis for a given data.

**COURSE OUTCOMES:**

Upon successful completion of this course, students will have the ability to:

| CO No. | Course Outcomes | Blooms Level | Module No. | Mapping with POs |
|---|---|---|---|---|
| 1 | Familiarize with foundations of Human learning and Machine learning and Basic types of data in Machine Learning applications. | Apply | 1 | 1, 2, 3, 4, 9, 12 |
| 2 | Apply probability and statistical methods to discrete and continuous variables. | Apply | 1 | 1, 2, 3, 4, 9, 12 |
| 3 | Apply the concepts in probability distributions and mathematical expectation inmachine learning applications. | Analyze | 2 | 1, 2, 3, 4, 9, 12 |
| 4 | Execute concept of correlation analysis and least square method in fitting regression curves to machine learning problems. | Analyze | 2 | 1, 2, 3, 4, 9, 12 |

**TEXT BOOKS:**

1. I.R. Miller, J.E. Freund and R. Johnson, "Probability and Statistics for Engineers", 9th Edition,Pearson, 2018.
2. Sheldon M. Ross, An Introduction to Probability and Statistics for Engineers and Scientists, 4th Edition, Academic Press, Elsevier, 2014.

**REFERENCE BOOKS:**

1. Kishore S. Trivedi, "Probability and Statistics with Realiability, Queueing and Computer Science Applications", 2nd edition, Wiley Student edition, 2008.
2. A Singaravelu, "Probability and Statistics", 22nd edition, Meenakshi Agency, 2015.